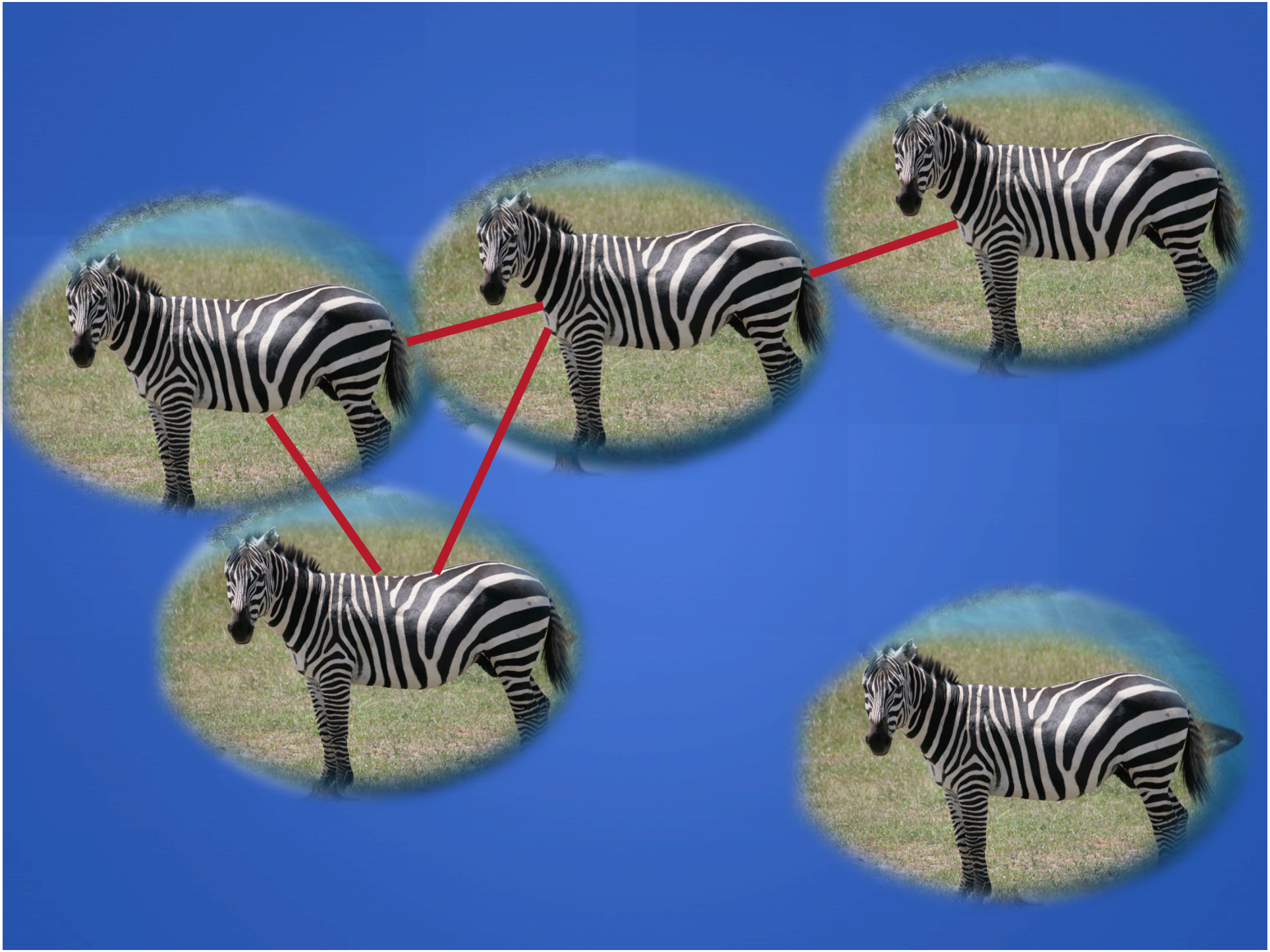
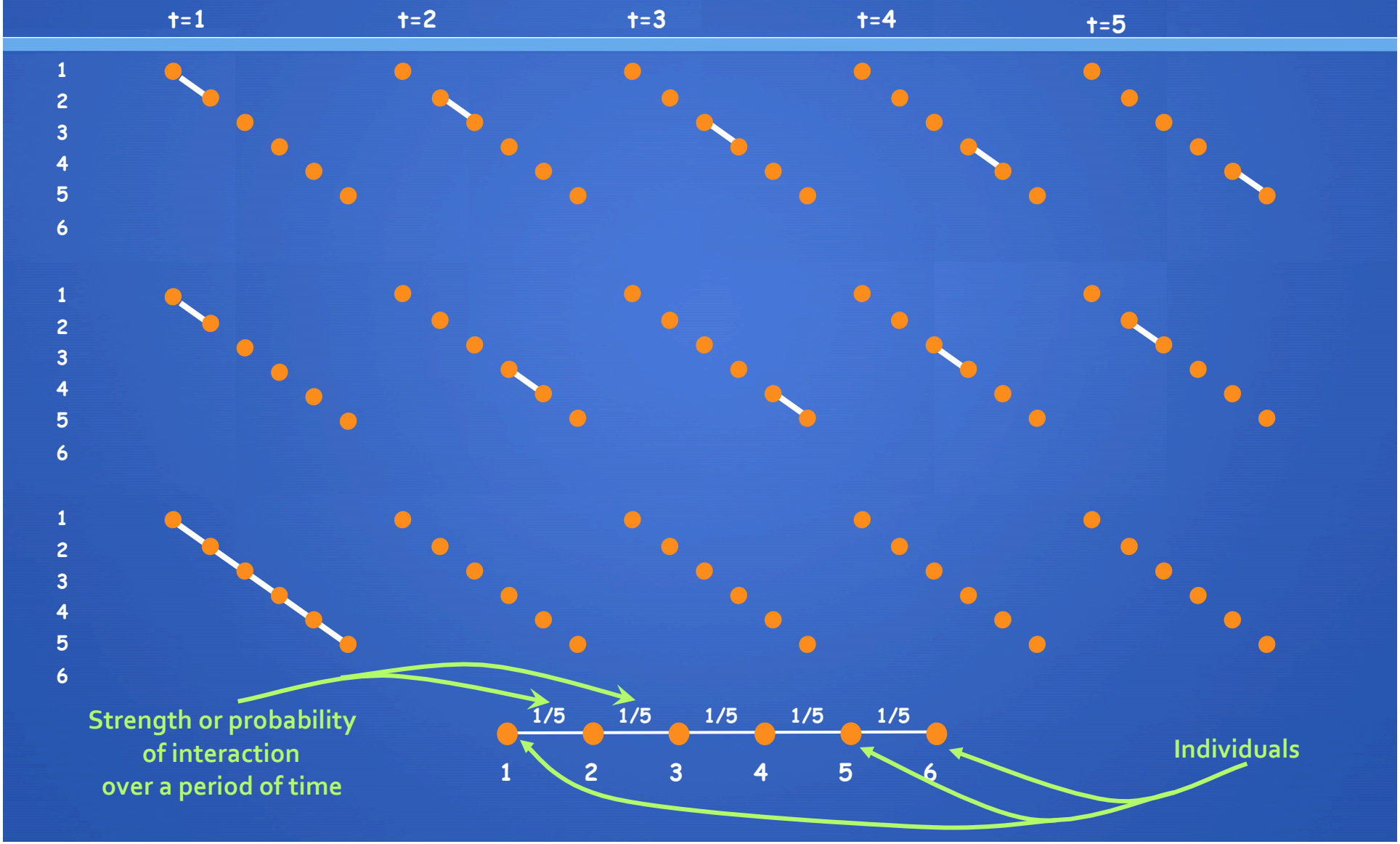
Computational Insights into Population Biology

Tanya Berger-Wolf

Computational Population Biology
University of Illinois at Chicago



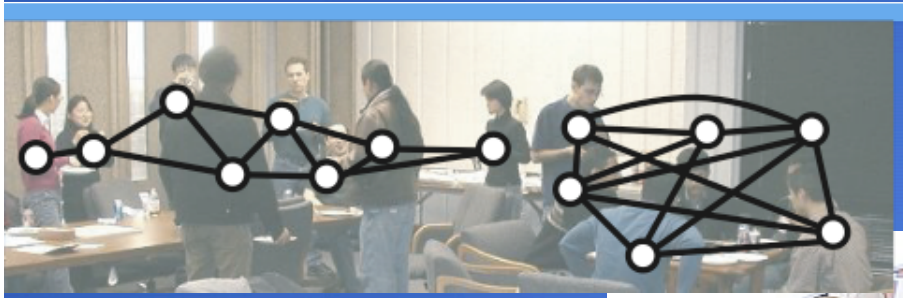
Social Networks: Static vs Dynamic

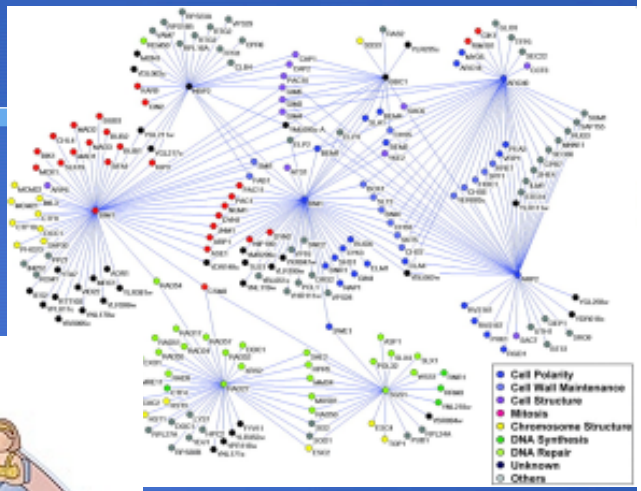


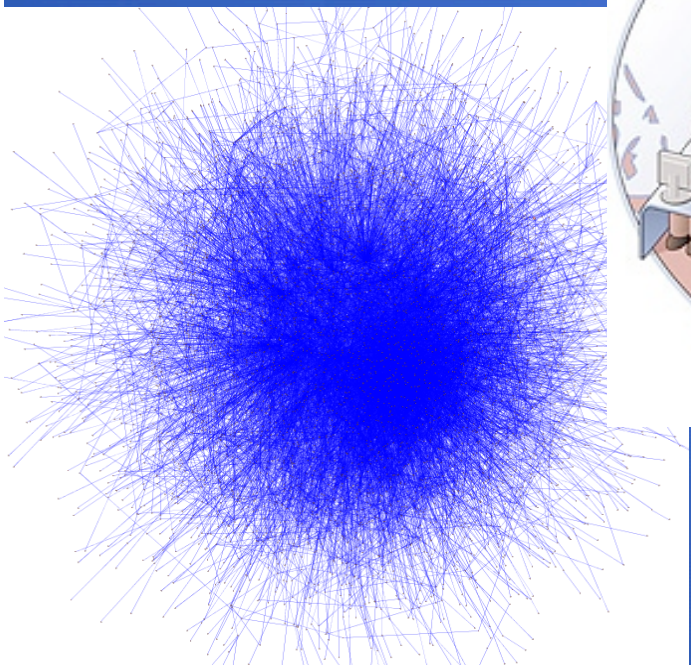
Advantage of Dynamic Networks:

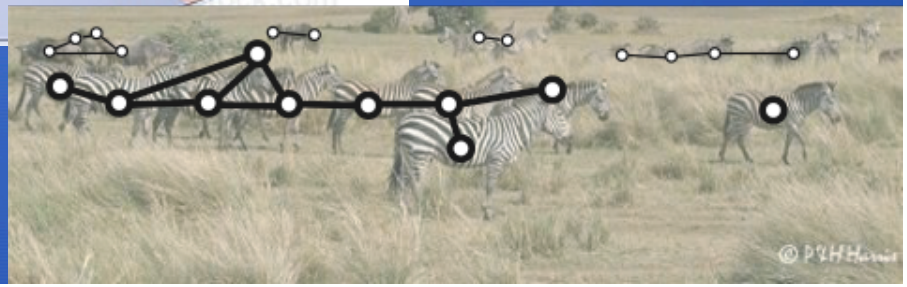
- More accurate information
- Time related questions:
  - How do diseases/information spread through population?
  - Who are the individuals that change the dynamics of interaction (leaders, interaction facilitators, etc.)? How do they emerge?
  - How do social structures (communities) change with outside circumstances ?
  - What is the average lifespan of a social structure and are there recurring structures?

Dynamics Networks

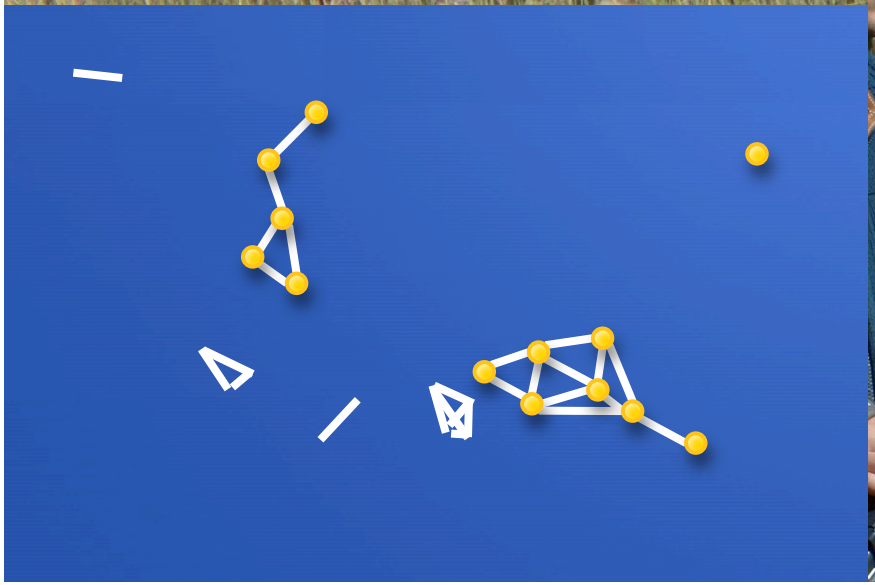


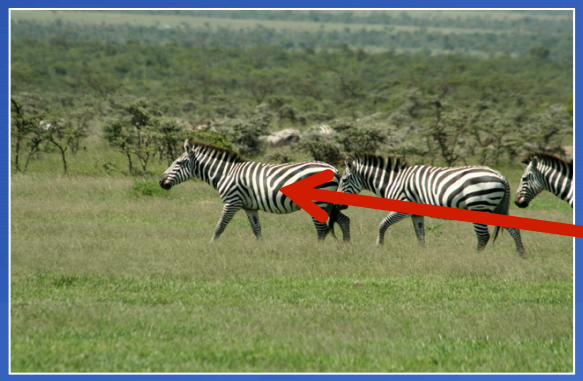




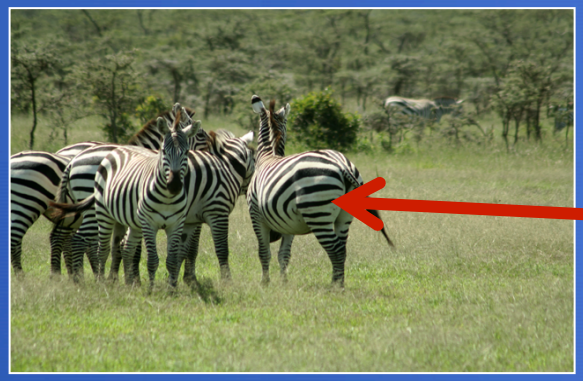


How do we get individual information?





"Moe"
April 23



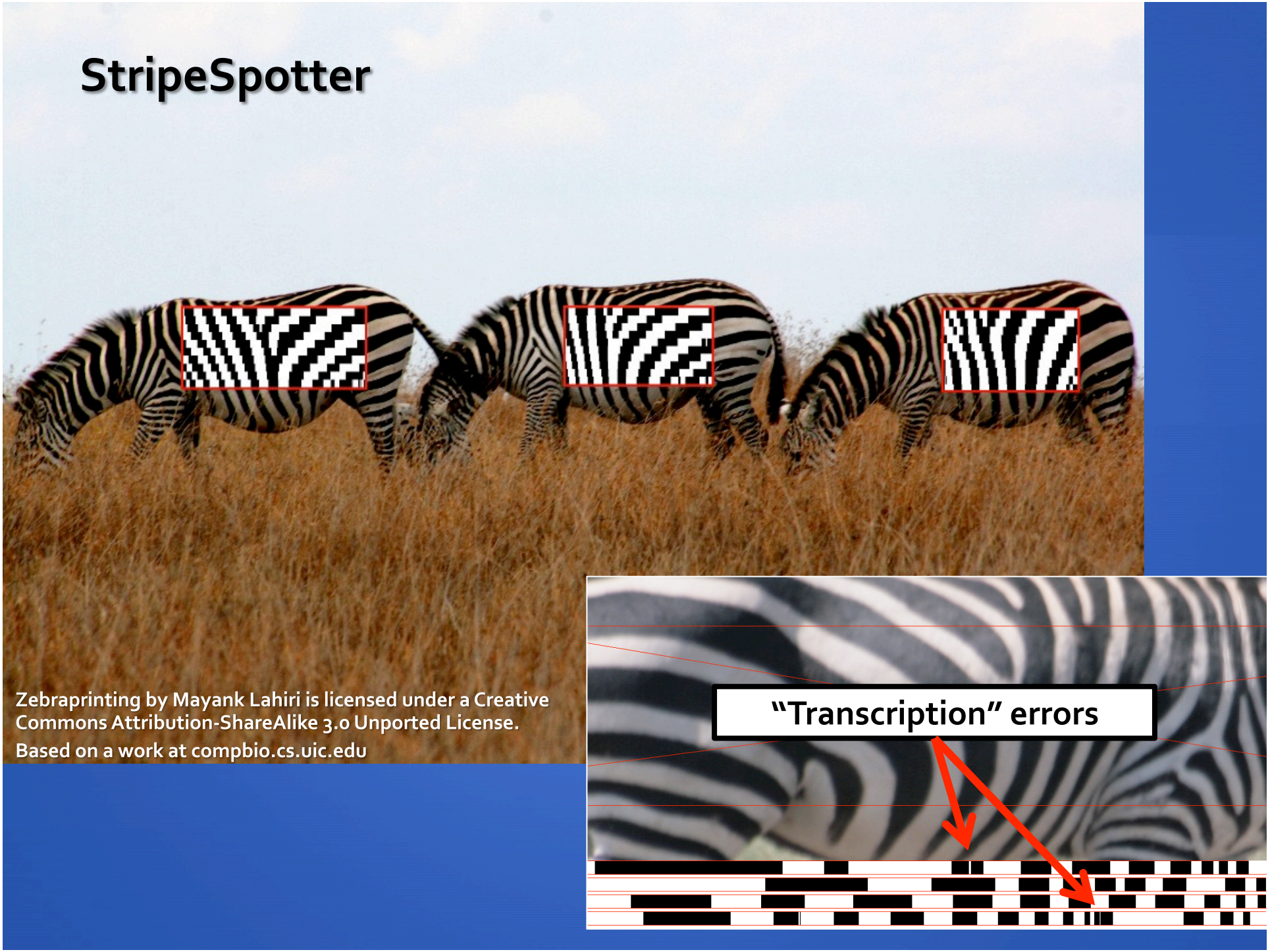
"Moe"
June 6

Telling zebras apart is a hard road
and I'm making mistakes by the car load.
So I got this device that keeps track of their stripes
and just scans them like they are a … BARCODE!

NPR, Wait Wait Don't Tell Me, April 16, 2011

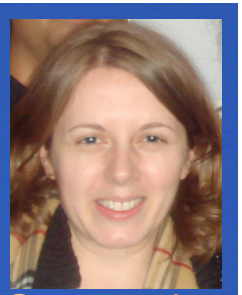
Image: Tim Flach/Getty

StripeSpotter

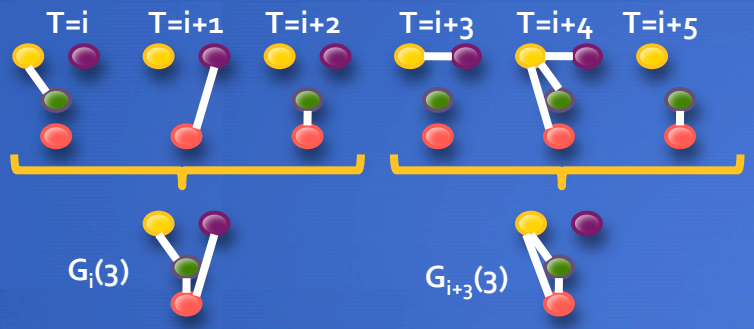



16
15
15

What is a time step?



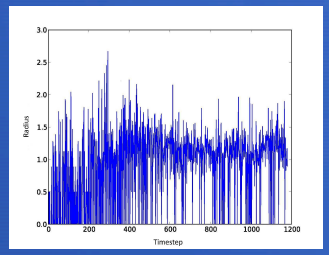
Sulo Caceres, B-W , Grossman '10

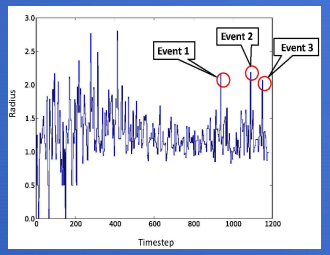


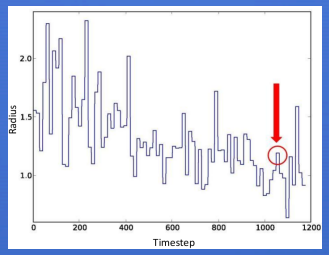
Temporal **W**indow **I**n **N**etworks Algorithm:

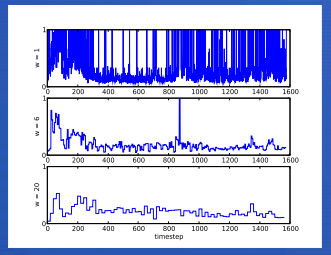
1. Given a graph sequence, compute the time series of the measure defined over a dynamic network with fixed resolution w.
2. Let **Variance** of the time series to represent noise. Let **Compression ratio** of the time series represent information.
3. Pick the window values for which Variance is low and compression ratio is high.

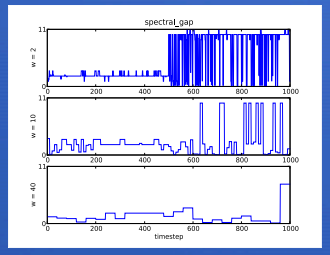
Any noisy stationary probabilistic process that generates the edge stream, which is oversampled at factor *w* will be stationary at aggregation window *w* and not less.

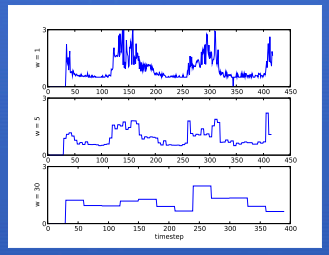


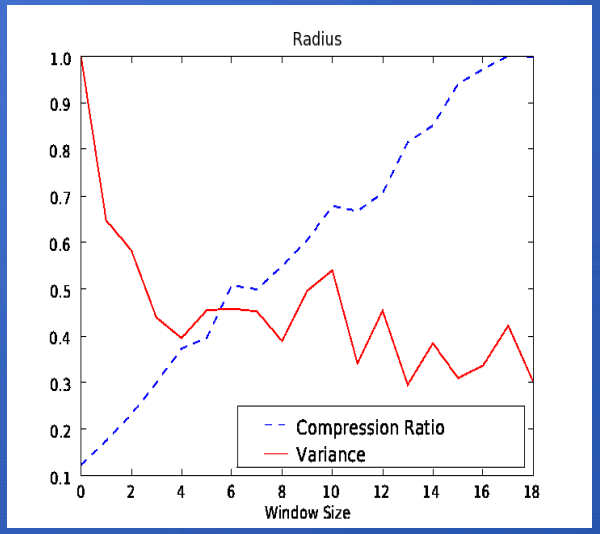






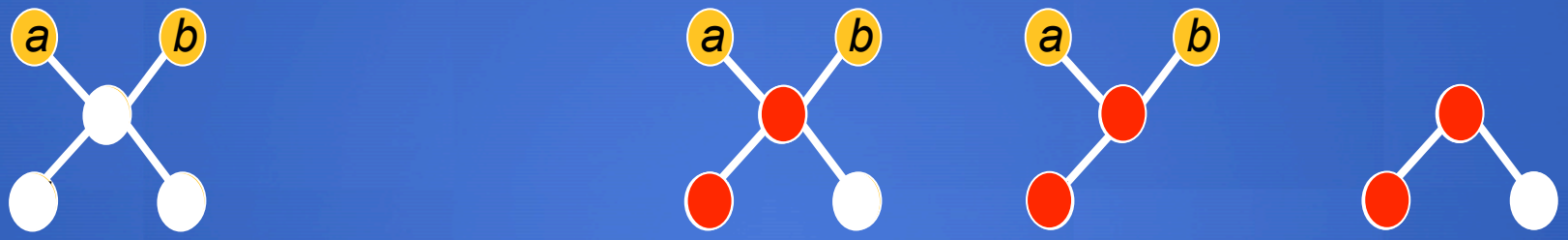






Periodic Pattern Discovery

Lahiri & B-W '08 + '10



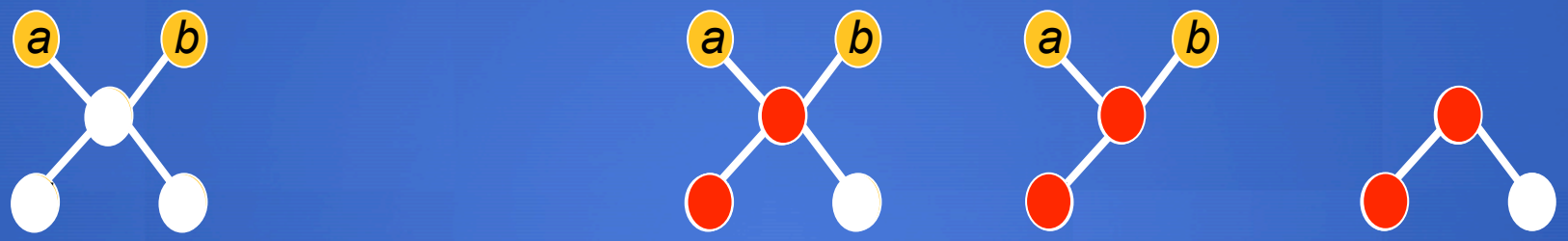
t = 1 *t = 2* *t = 3* *t = 4* *t = 5*

Given a dynamic network, find all subgraphs that occur periodically more than σ times.

Downward closure: Any subgraph of a periodic subgraph is also periodic.

Periodic Pattern Discovery

Lahiri & B-W '08 + '10



Given a dynamic network, find all subgraphs that occur periodically more than σ times.

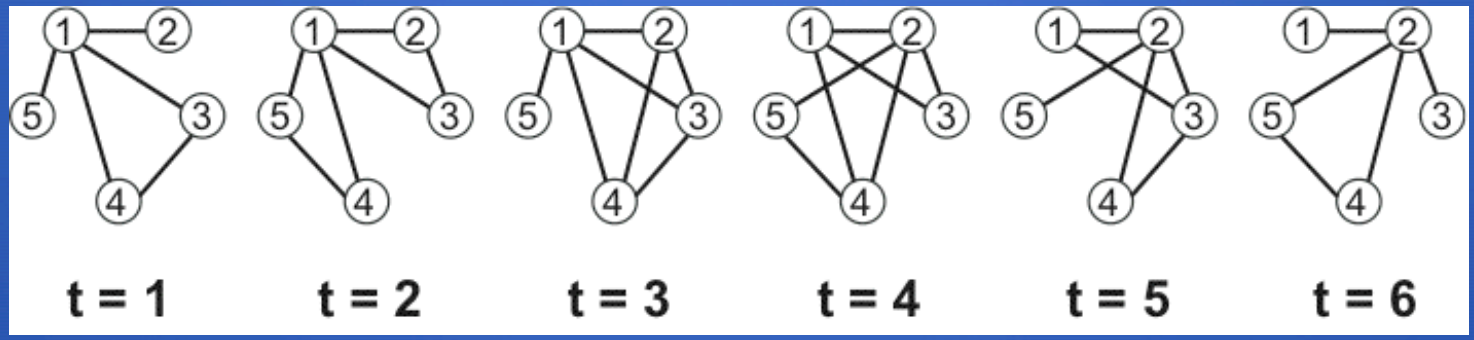
1. Occurs more than σ times
2. Cannot be extended structurally
3. Cannot be extended in time

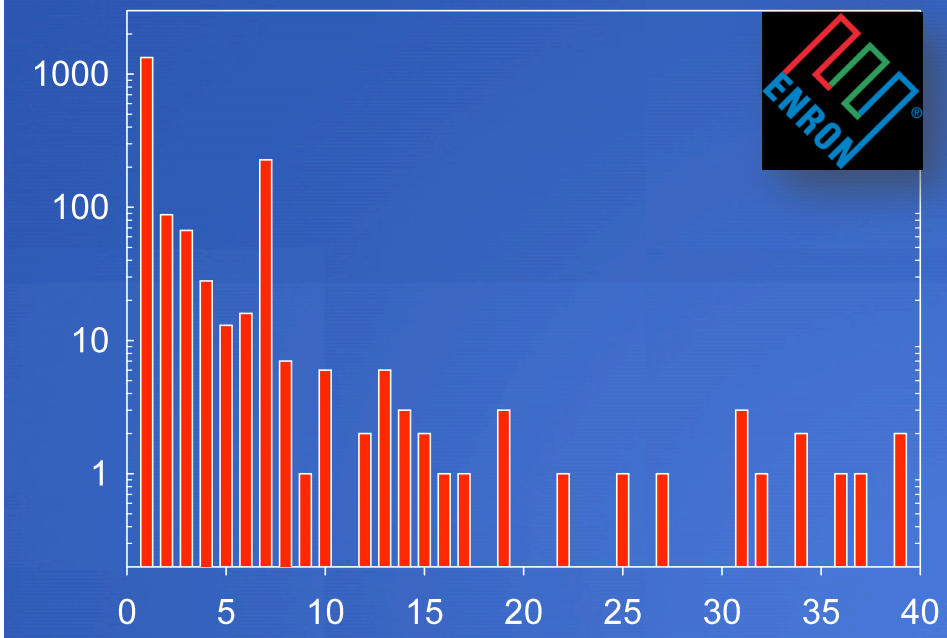
} Closed periodic subgraphs

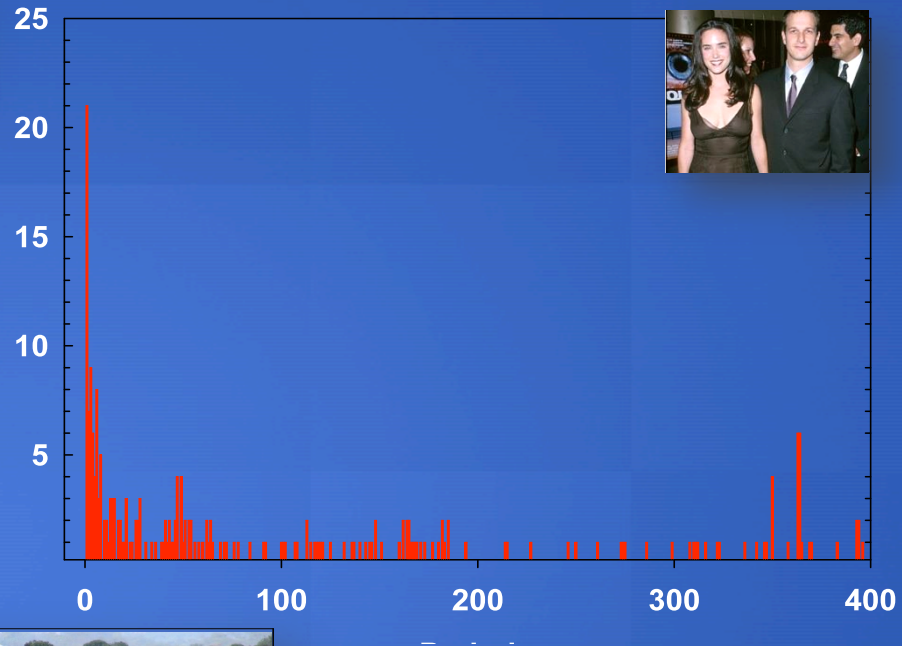
Periodic Pattern Discovery

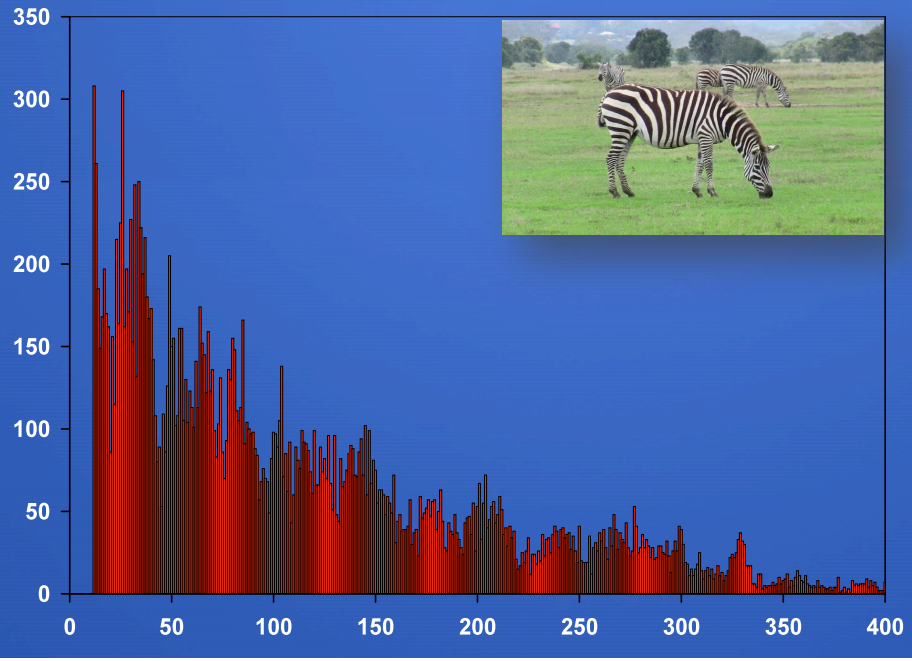
- In a finite amount of data:
  - Maximum period has a natural upper bound
- Construct worst-case example from a mining point of view
  - Each possible periodic subgraph is present
  - Iterate over all possible periodic positions, "creating" a new subgraph
  - Polynomial number of closed periodic subgraphs.
- Algorithm to enumerate this is $O(T^2 \ln(T/\sigma) (V+E))$*

  * When each graph has unique node labels




1000
100
10
1
0
5
10
15
20
25
30
35
40

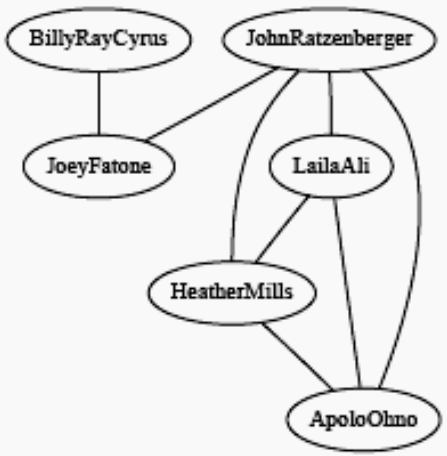

25
20
15
10
5
0
100
200
300
400


350
300
250
200
150
100
50
0
0
50
100
150
200
250
300
350
400

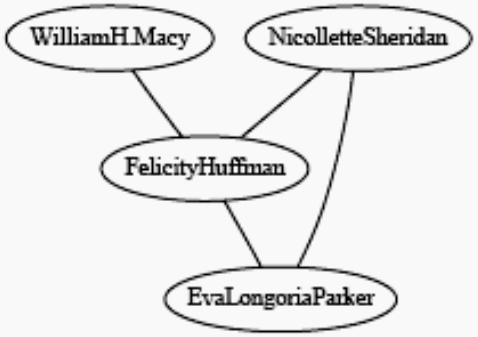
IMDB.com

IMDB.com

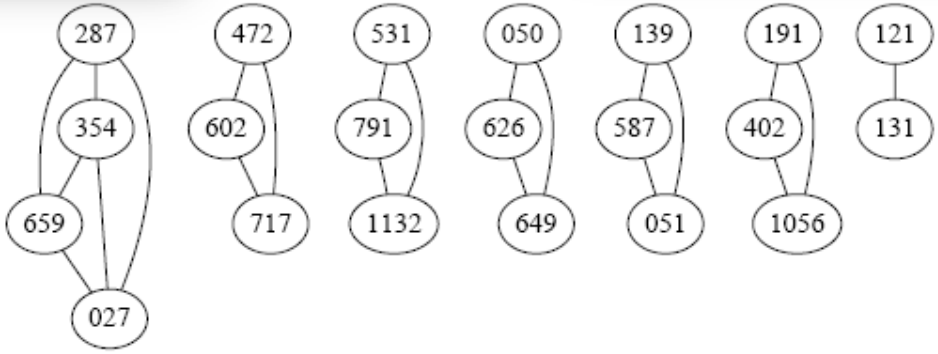
IMDB.com



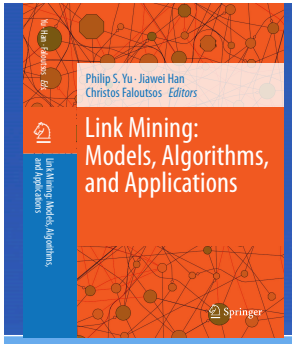
(a) IMDB: period 7 ± 2, support 3, avg. purity 1



(b) IMDB: period 364, support 3, avg. purity 0.4

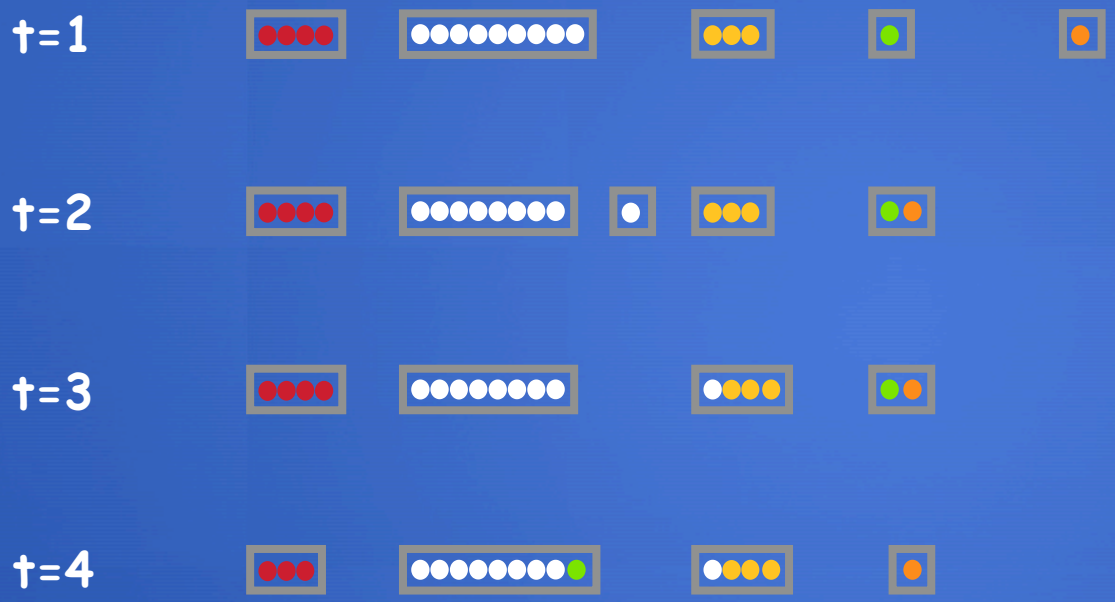


(d) Plains: period 7, support 4, avg. purity 0.94.



Dynamic Communities

Tantipathananandh, TB-W & Kempe '07
Tantipathananandh, TB-W '09



Individuals change affiliations, group split and merge

Are there cohesive groups?

How does community structure arise from interactions and how are the interactions affected by the latent community structure?

Want to allow fluid community structure

Dynamic Communities

What is a community?

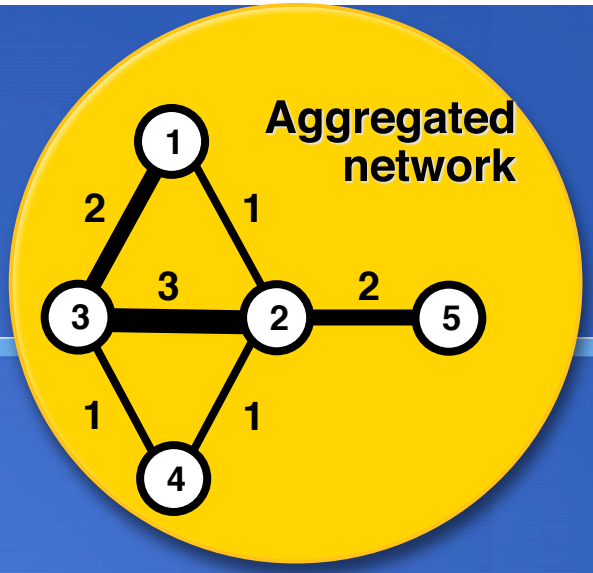
"Cohesive subgroups are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties." [Wasserman & Faust '97]

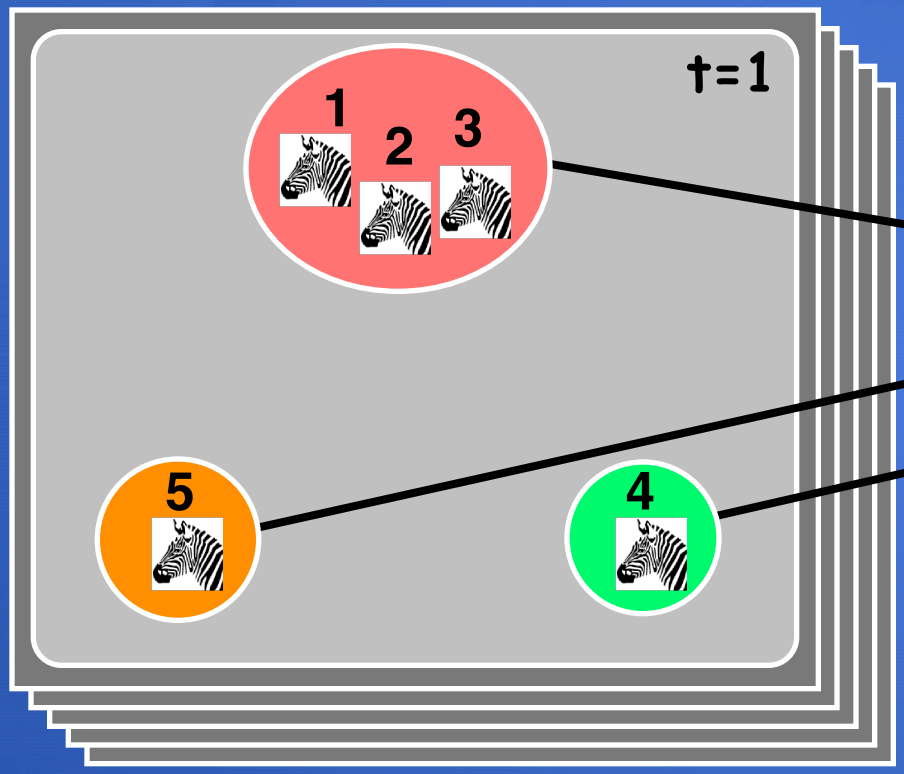
Community = cluster
Dynamic community = dynamic cluster: members of community interact more frequently (consecutively) among them than with individuals outside the community.
Dynamic community = identity.

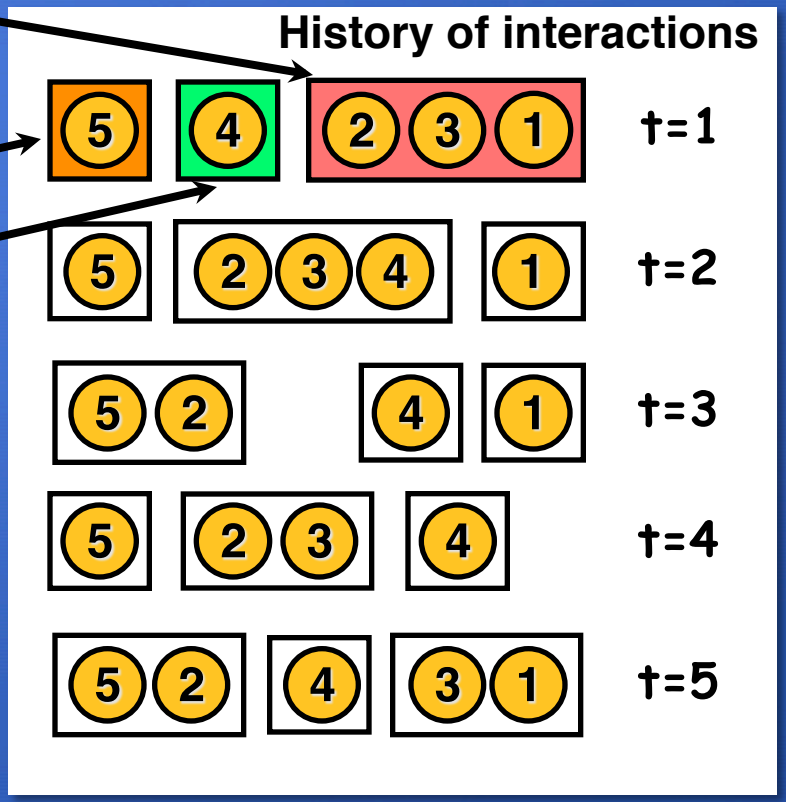
Computational Approaches

- Aggarwal and Yu '05
- Metagroups [Berger-Wolf & Saia '06]
- MONIC [Spiliopoulou et al '06]
- Palla et al '07
- Falkowski et al '06
- GraphScope [Sun et al '07]
- FacetNet [Lin et al '08], Yang et al '09, Colibri [Tong et al '08], Sarkar and Moore '05, Fu et al '09

Formal Representation

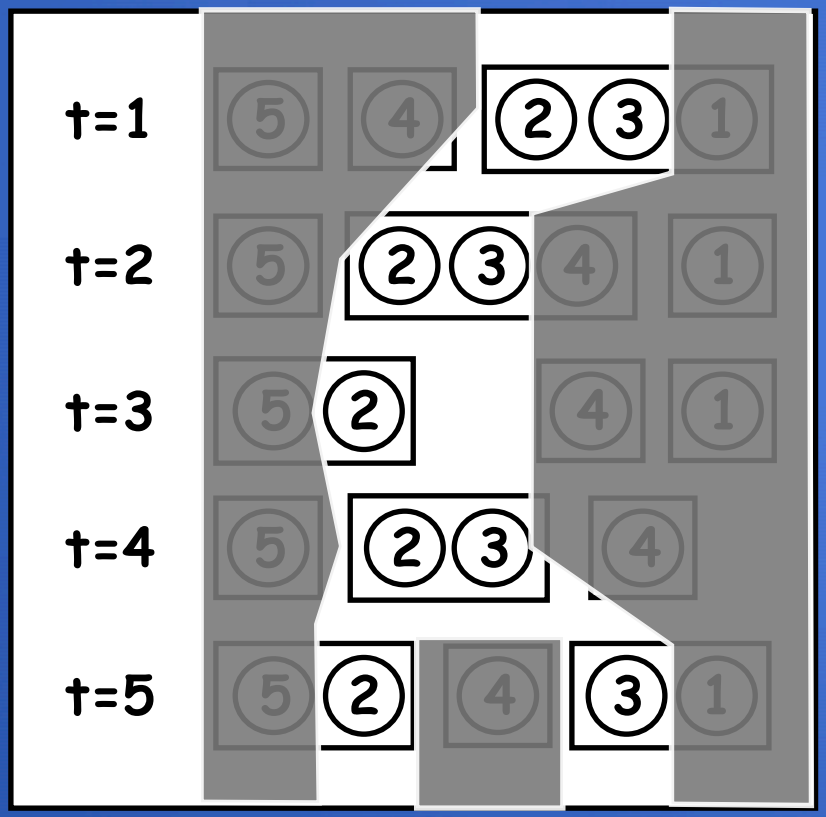






Assume discrete time and interactions in form of complete subgraphs.

The Question: What is dynamic community?



- A dynamic community is a subset of individuals that stick together over time.
- Community is a latent object
- Community = object's identity
- NOTE: Communities ≠ Groups

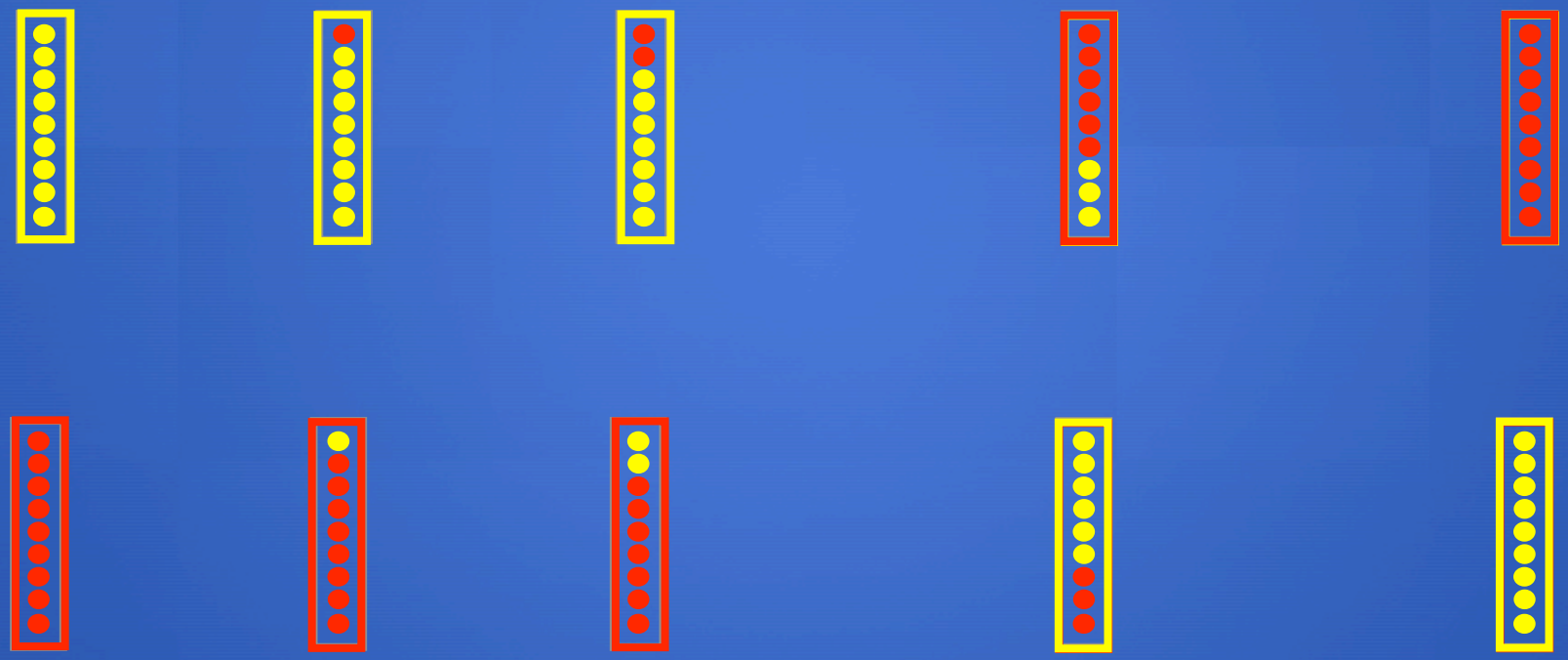
Theseus's Paradox

- During a twelve month period 95% of all the atoms that make up your 50 trillion cells are replaced
- FAA regulations: airplane = left rudder number
- **Ship of Theseus**

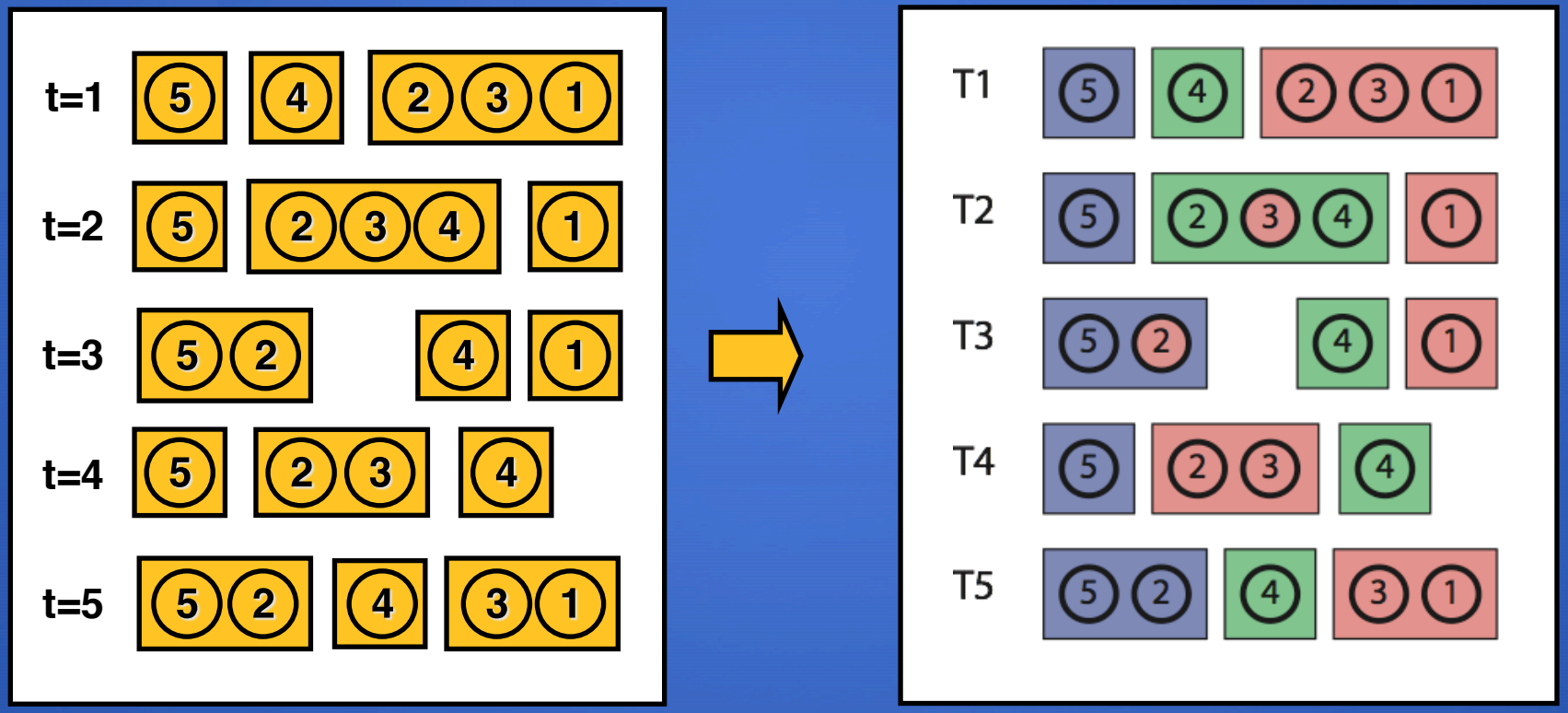
  "The ship wherein Theseus and the youth of Athens returned [from Crete] had thirty oars, and was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting in new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same."

Theseus' Ship as Dynamic Community

t=1 t=2 t=3 … t=7 … t=10



Community = Color



Assumptions

- Individuals are reluctant to switch community affiliations
  cost of switching
- Individuals mostly are seen with their own community
  cost of visit
- Individuals a rarely absent from their own community
  cost of absence

Parsimony = minimize the total cost across all individuals

Community as a Coloring

Minimum Community Interpretation:

For a given cost setting, $(\alpha, \beta_1, \beta_2)$, find vertex coloring that minimizes total cost.

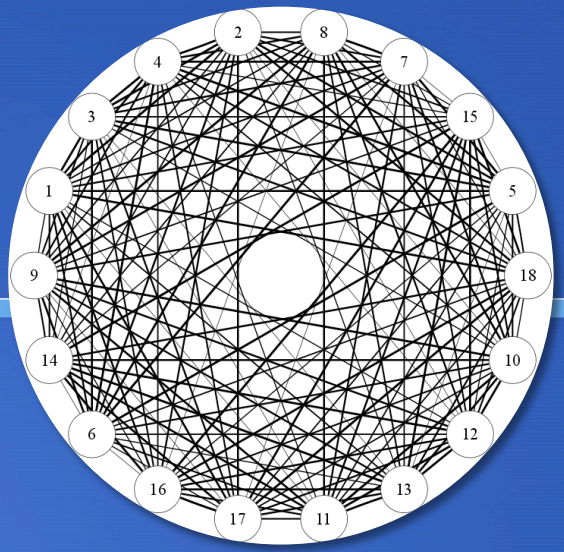
Color of group vertices = Community structure

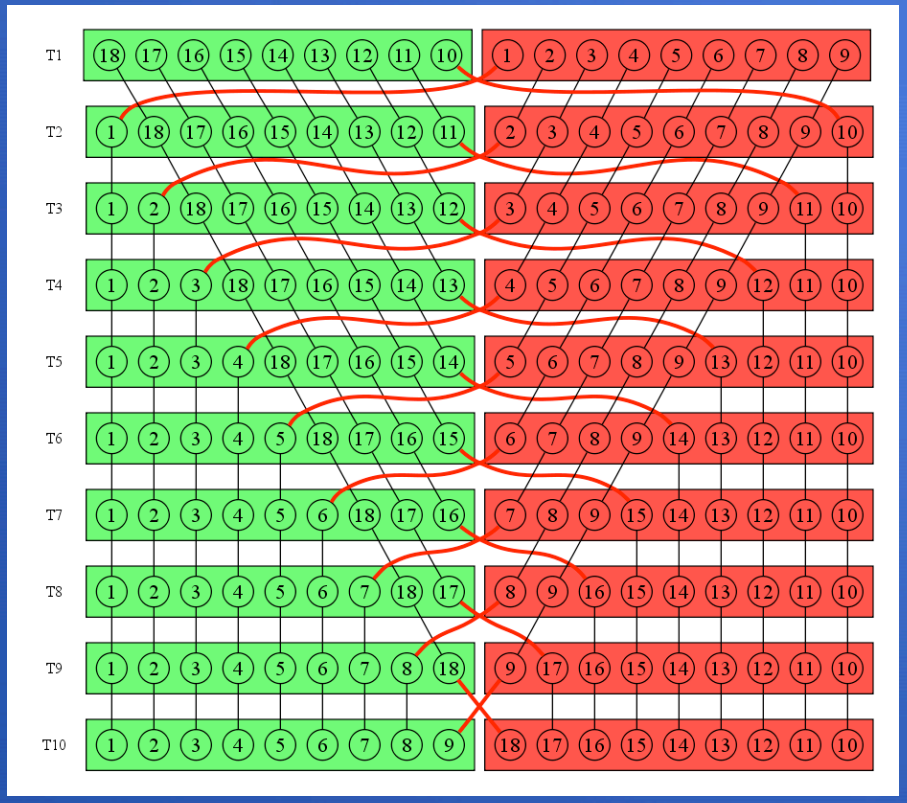
Color of individual vertices = Affiliation sequences

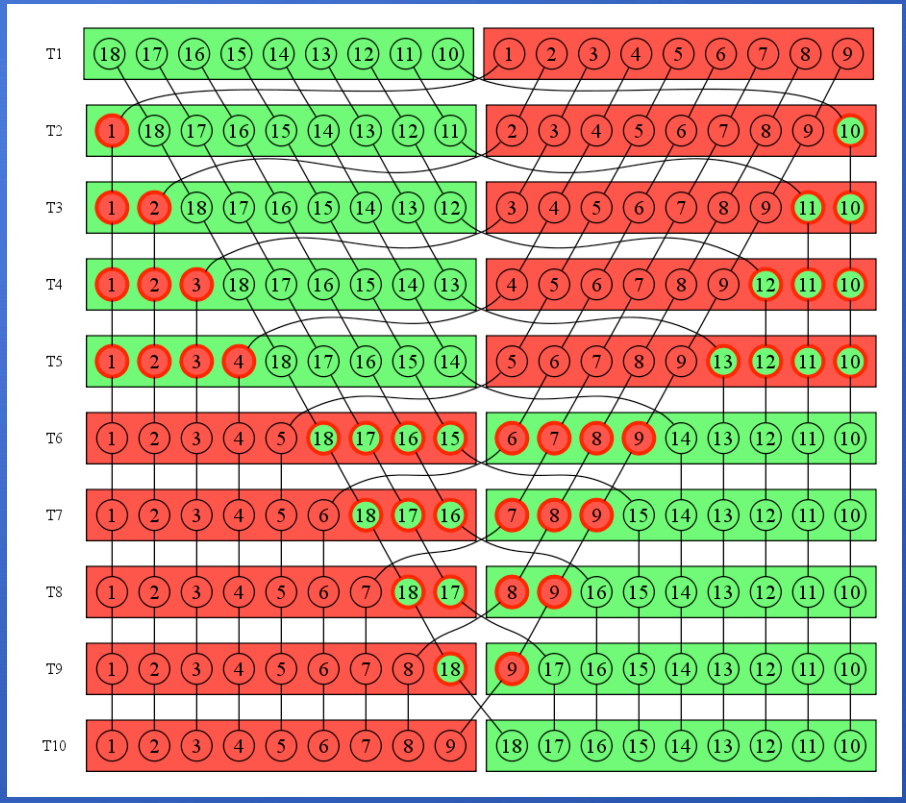
Problem is NP-Complete and APX-Hard

But fast flow-based constant factor approximations

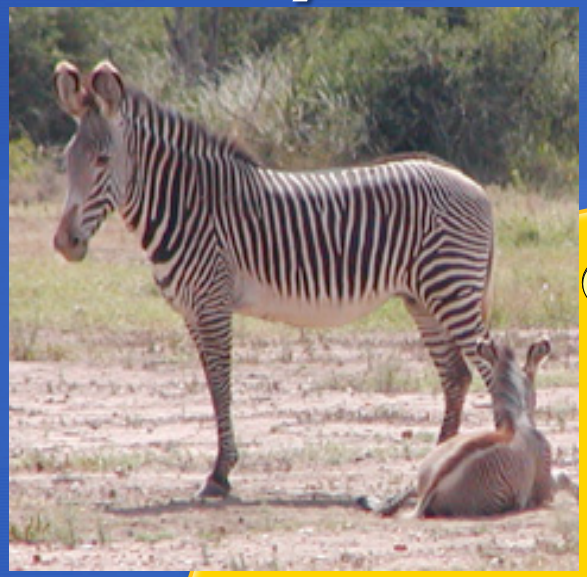
Theseus' Ship

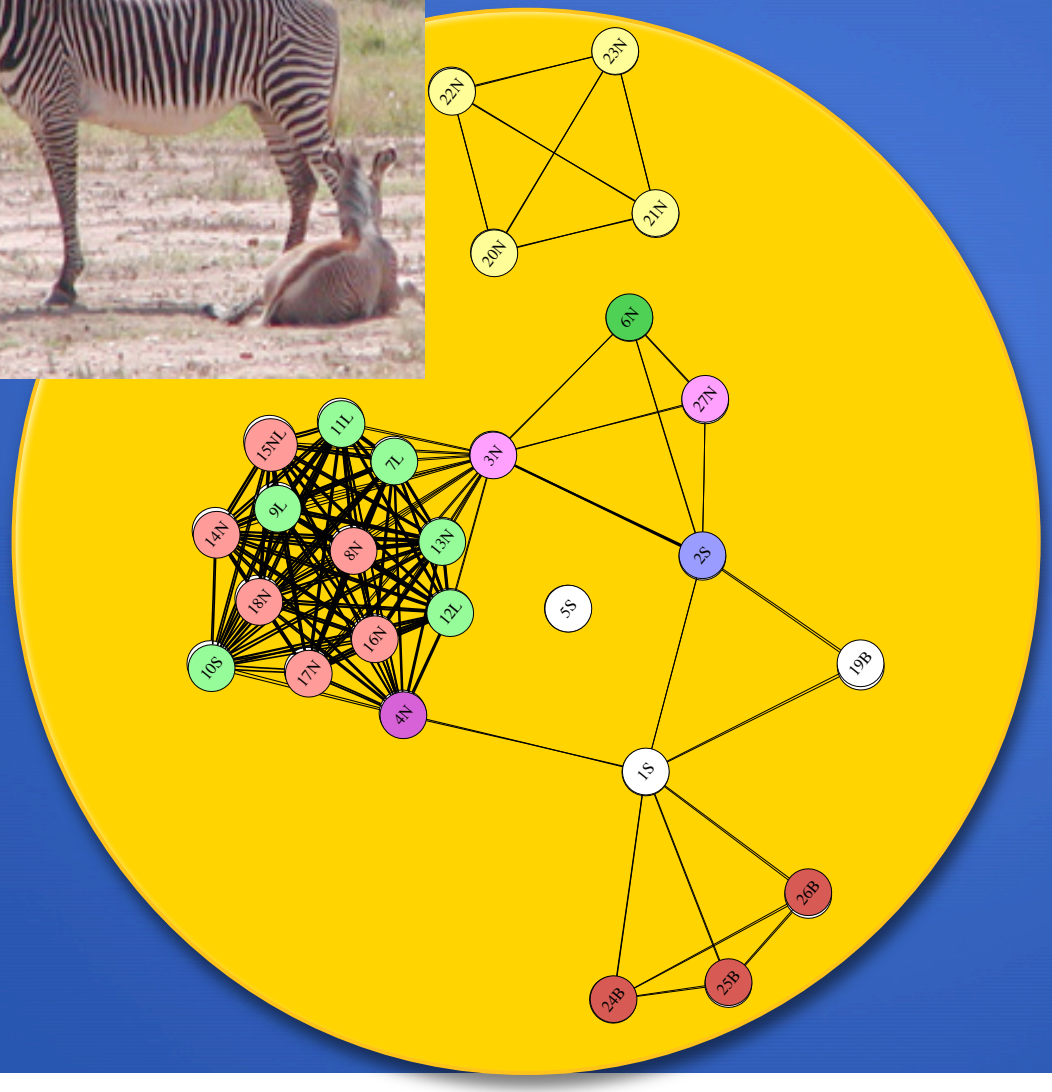


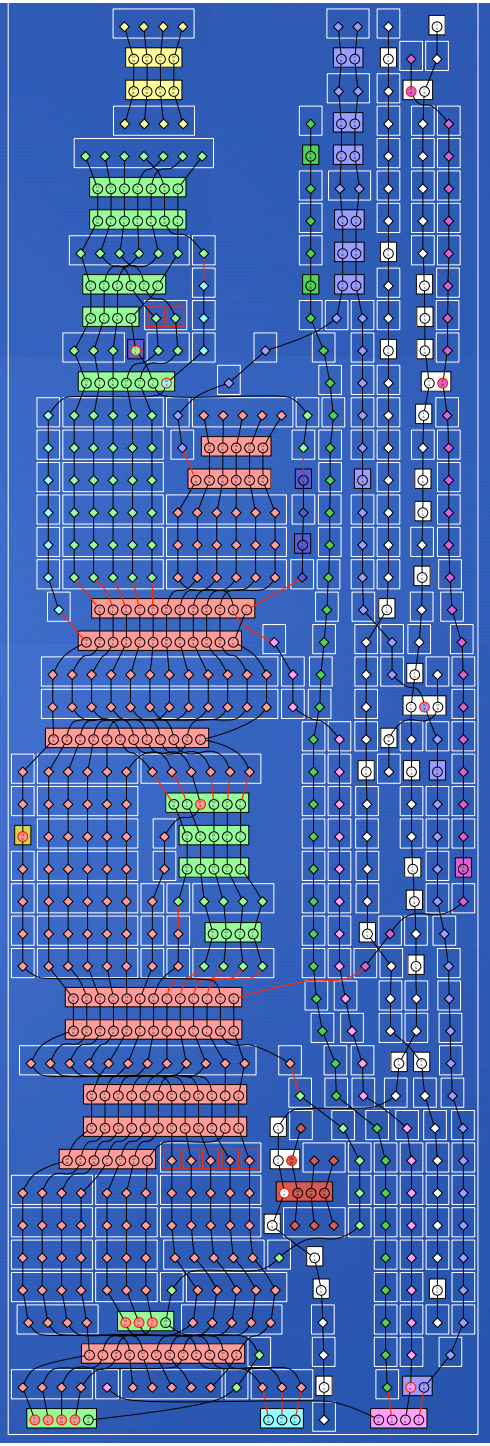


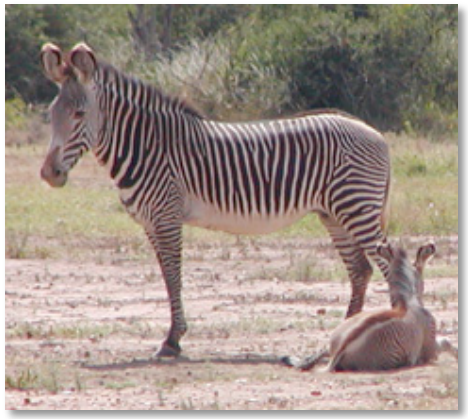


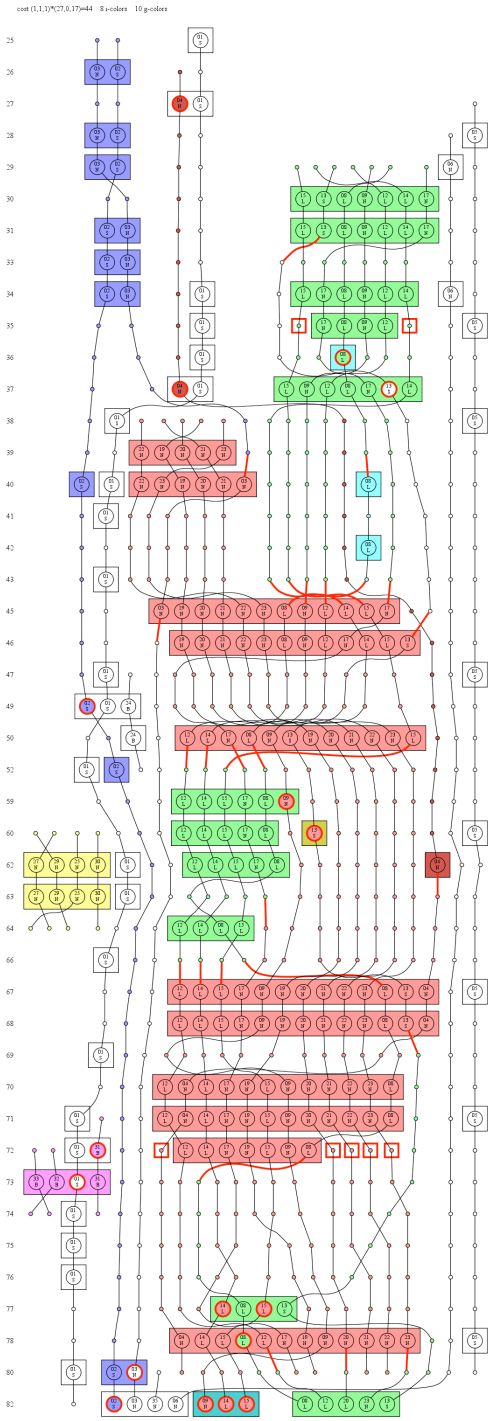
Grevy's Zebra

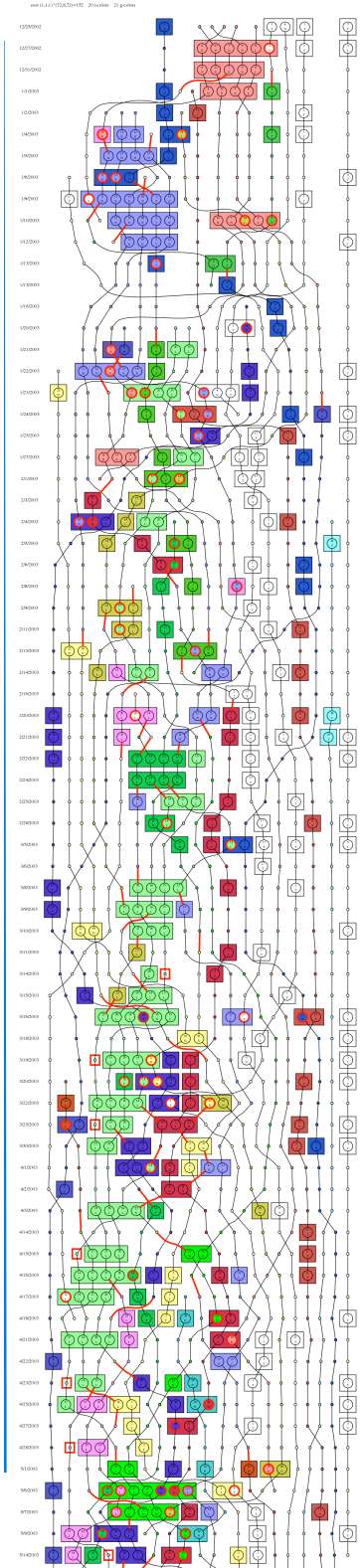


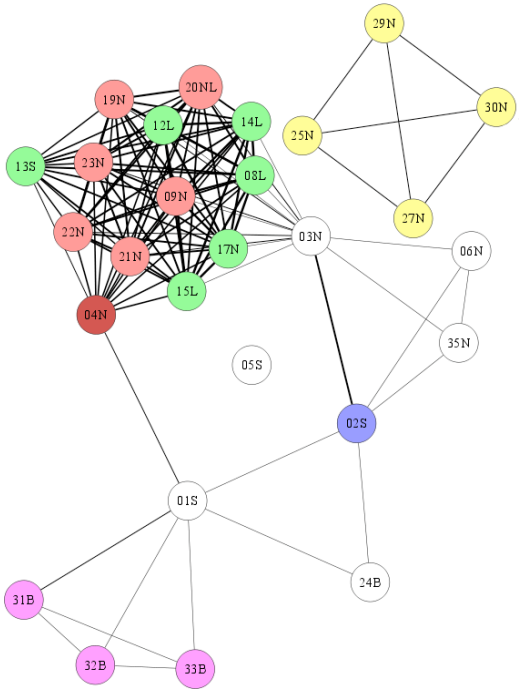


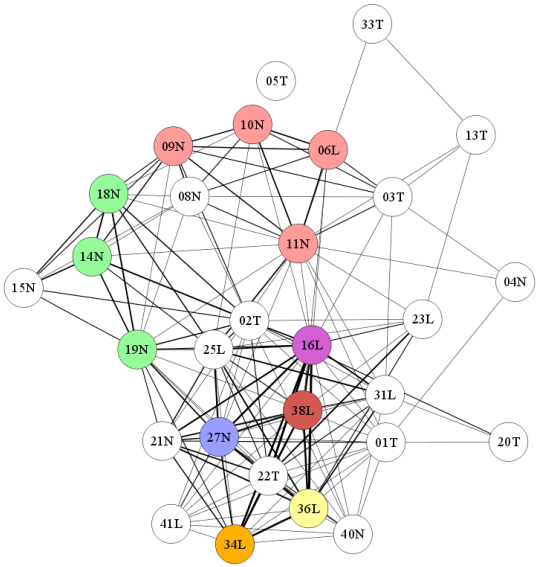





cost (1,1,1)*(27,0,17)=44 8 i-colors 10 g-colors

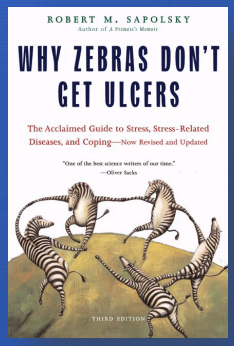



29N
30N
25N
27N
19N
20NL
12L
14L
13S
23N
08L
09N
22N
21N
17N
03N
06N
15L
04N
35N
05S
02S
01S
24B
31B
32B
33B

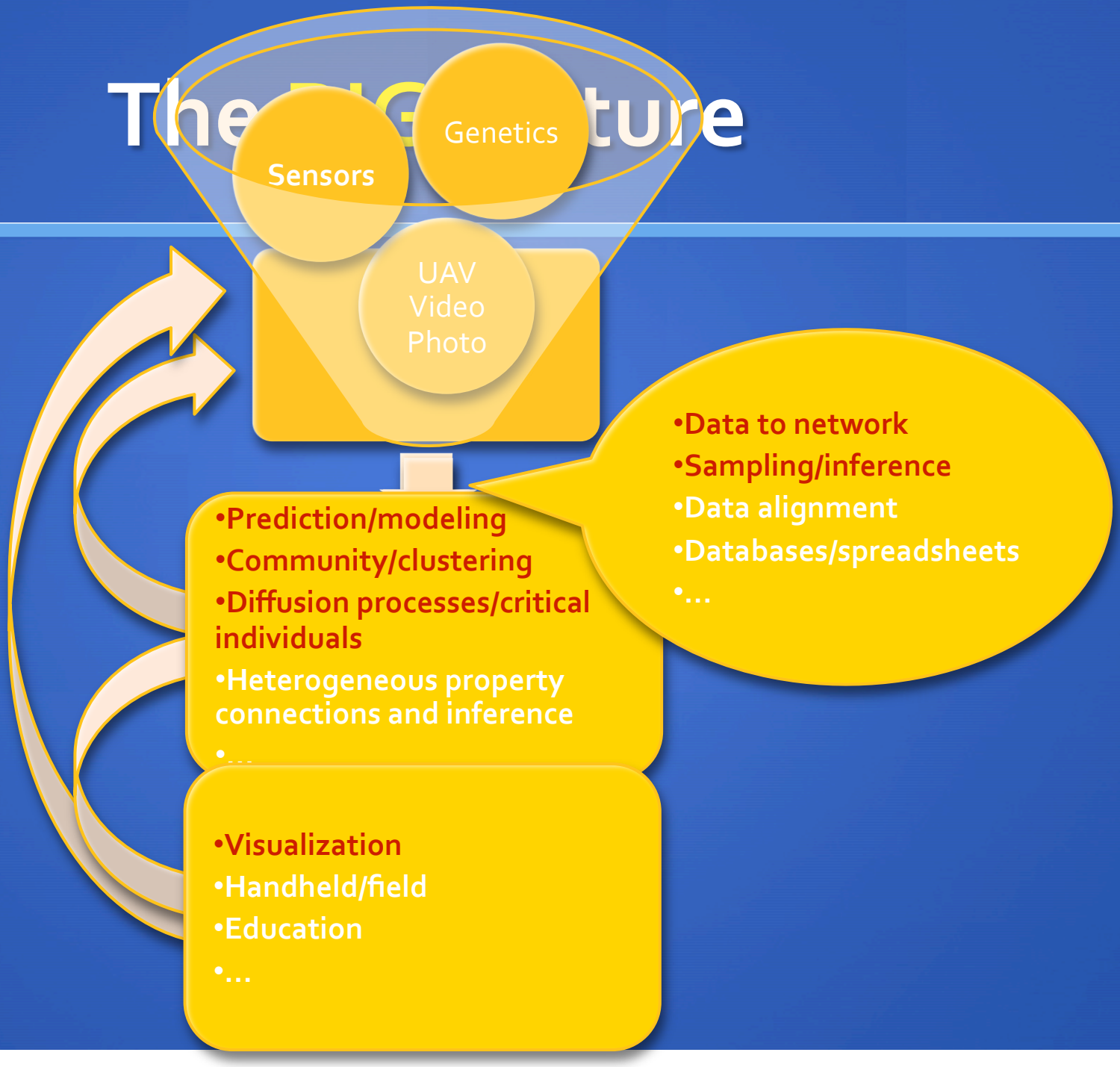

33T
05T
10N
13T
09N
06L
18N
08N
03T
14N
11N
15N
04N
02T
23L
19N
25L
16L
31L
38L
21N
27N
01T
20T
22T
36L
41L
34L
40N

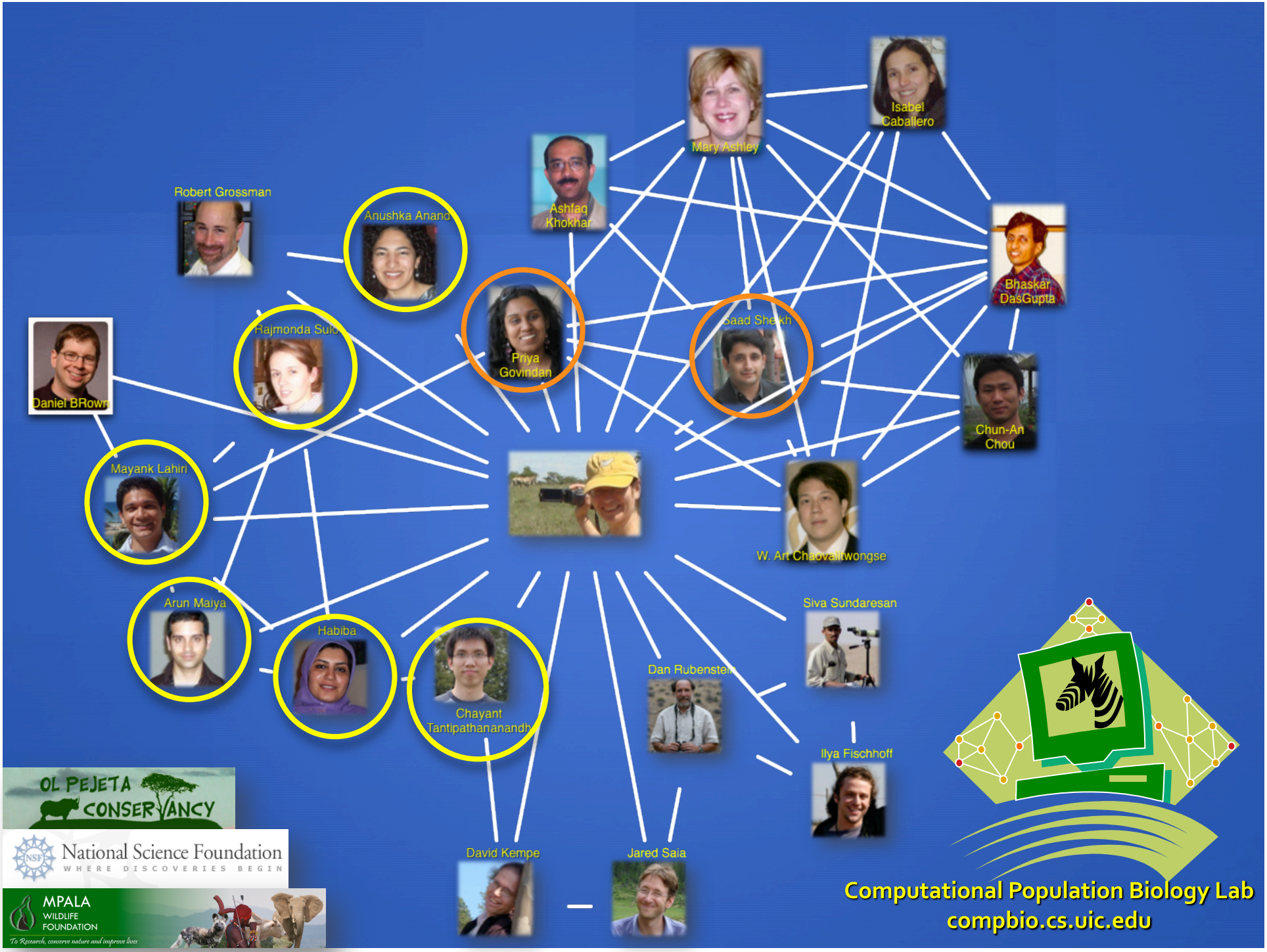
So

- Dynamic is really different from static
- Must develop computational methods that address the order, concurrency, and delay of interactions
- We have a few things but this is just starting – lots more work to do!
  (prediction, communities and other structures, hierarchies, spread, sampling, ….)
- Zebras are better than humans
  (they don't have ulcers—apologies to R. Sapolski)
  - No cellphones or emails: proximity ≈ interaction
  - No privacy issues
  - Can conduct experiments to validate and test
- But everything we do applies to humans



The BIG Picture




Mary Ashley
Isabel Caballero
Ashfaq Khokhar
Robert Grossman
Anushka Anand
Bhaskar DasGupta
Priya Govindan
Saad Sheikh
Daniel BRown
Chun-An Chou
Mayank Lahiri
W. Art Chaovalitwongse
Arun Maiya
Habiba
Siva Sundaresan
Chayant Tantipathananandh
Dan Rubenstein
Ilya Fischhoff
David Kempe
Jared Saia
OL PEJETA CONSERVANCY
National Science Foundation
WHERE DISCOVERIES BEGIN
MPALA WILDLIFE FOUNDATION
To Research, conserve nature and improve lives
Computational Population Biology Lab
compbio.cs.uic.edu